ANALISIS REGRESI
untuk Penelitian

Prof. Dr. Suyono, M.Si

Analisis Regresi untuk Penelitian

UU No 19 Tahun 2002 Tentang Hak Cipta

Fungsi dan Sifat hak Cipta Pasal 2

1. Hak Cipta merupakan hak eksklusif bagi pencipta atau pemegang Hak Cipta untuk mengumumkan atau memperbanyak ciptaannya, yang timbul secara otomatis setelah suatu ciptaan dilahirkan tanpa mengurangi pembatasan menurut peraturan perundang-undangan yang berlaku.

Hak Terkait Pasal 49

1. Pelaku memiliki hak eksklusif untuk memberikan izin atau melarang pihak lain yang tanpa persetujuannya membuat, memperbanyak, atau menyiarkan rekaman suara dan/atau gambar pertunjukannya.

Sanksi Pelanggaran Pasal 72

1. Barangsiapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal 2 ayat (1) atau pasal 49 ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp 1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp 5.000.000.000,00 (lima miliar rupiah).
2. Barangsiapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran Hak Cipta sebagaimana dimaksud dalam ayat (1), dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp 500.000.000,00 (lima ratus juta rupiah)

Analisis Regresi untuk Penelitian

Prof. Dr. Suyono, M.Si.

Jl. Rajawali, G. Elang 6, No 3, Drono, Sardonoharjo, Ngaglik, Sleman
Jl.Kaliurang Km.9,3 – Yogyakarta 55581
Telp/Faks: (0274) 4533427
Website: www.deepublish.co.id
www.penerbitdeepublish.com
e-mail: deepublish@ymail.com

Katalog Dalam Terbitan (KDT)

SUYONO
Analisis Regresi untuk Penelitian /oleh Suyono.--Ed.1, Cet. 1--
Yogyakarta: Deepublish, Agustus 2015.

xi, 294 hlm.; Uk:15.5x23 cm

ISBN **978-602-401-025-6**

1. Analisis Regresi I. Judul
519.53

Desain cover : Herlambang Rahmadhani
Penata letak : Cinthia Morris Sartono

PENERBIT DEEPUBLISH
(Grup Penerbitan CV BUDI UTAMA)
Anggota IKAPI (076/DIY/2012)

Kata Pengantar

Bismillahirrohmaanirrahiim.

Puji syukur kami panjatkan kepada Allah Swt. karena hanya atas nikmat dan karunia-Nya buku ini dapat kami selesaikan. Sholawat dan salam semoga tercurah atas Nabi Muhammad Saw.

Buku ini kami beri judul *Analisis Regresi untuk Penelitian* karena dimaksudkan untuk membantu para pembaca dalam menganalisis data hasil penelitian dengan menggunakan analisis regresi. Buku ini tidak hanya menjelaskan bagaimana menggunakan rumus-rumus yang ada secara tepat, tetapi juga menyajikan konsep-konsep penting dalam analisis regresi yang dapat dipahami oleh pembaca tanpa harus memiliki pengetahuan matematika yang mendalam. Dengan demikian, diharapkan para pembaca dapat menyimpulkan dan menginterpretasikan hasil-hasil analisis data dengan benar. Untuk mempermudah dalam memahami buku ini contoh-contoh perhitungan analisis data disajikan, baik secara manual maupun dengan menggunakan *software* SPSS. Di dalam buku ini juga dibahas analisis regresi dalam notasi matriks yang memiliki banyak keuntungan dengan disertai contoh-contoh yang perhitungannya dapat dilakukan dengan sangat mudah dengan menggunakan *software* Microsoft Excel.

Buku ini akan sangat bermanfaat bagi para mahasiswa S1, S2, dan S3 dari berbagai latar belakang bidang keilmuan dalam menganalisis data hasil penelitian dalam rangka menulis skripsi, tesis, atau desertasi. Buku ini juga dapat dimanfaatkan oleh para peneliti, praktisi, atau sebagai salah satu referensi perkuliahan.

Kami menyadari bahwa buku ini jauh dari sempurna. Oleh karena itu, koreksi, kritik, dan saran sangat kami harapkan dari para pembaca sehingga buku ini ke depan akan semakin baik lagi.

Akhirnya, melalui kesempatan ini kami ingin menyampaikan ucapan terima kasih kepada istri dan anak-anak yang telah memberi dukungan penuh sehingga buku ini akhirnya dapat kami selesaikan. Kami juga mengucapkan terima kasih kepada para mahasiswa Program Pascasarjana Universitas Negeri Jakarta yang telah memberi masukan selama draft buku ini digunakan sebagai bahan kuliah. Terima kasih juga kami sampaikan kepada Agus Agung Permana, S.Si yang telah membantu menyiapkan gambar-gambar di dalam buku ini. Semoga semuanya mendapat balasan yang baik dari Allah Swt. Amiin.

Jakarta, Agustus 2015

Penulis

Daftar Isi

Bab 1
Pendahuluan

1.1 Model Probabilistik

Lama waktu yang dibutuhkan oleh seorang mahasiswa untuk melakukan perjalanan dari rumah ke kampus dengan menggunakan kendaraan bermotor umumnya tergantung pada jarak dari rumah ke kampusnya. Anggap Anda akan memodelkan waktu tempuh sebagai fungsi dari jarak. Apakah Anda dapat memastikan berapa lama waktu tempuh yang diperlukan oleh seorang mahasiswa jika diketahui jarak dari rumah ke kampusnya? Anda semua mungkin setuju bahwa jawabannya 'tidak' karena ada faktor-faktor yang dapat memengaruhi lama waktu tempuh, misalnya berapa banyak persimpangan jalan yang dilalui, pukul berapa berangkat dari rumah, dan lain-lain.

Dalam 'kondisi ideal', jika seorang mahasiswa dapat melakukan perjalanan dengan kecepatan konstan 40 km/jam dan jarak dari rumah ke kampusnya 20 km, maka ia memerlukan waktu *tepat* setengah jam untuk melakukan perjalanan dari rumah ke kampus. Secara umum, jika seorang mahasiswa melakukan perjalanan dengan kecepatan konstan 40 km/jam dan jarak dari rumah ke kampus adalah X km maka waktu yang diperlukan (dalam satuan jam) adalah:

$$Y = \frac{1}{40} X \quad \text{atau } Y = 0{,}025\, X \qquad (1.1)$$

Model di atas merupakan **model deterministik** yang menyatakan hubungan antara jarak dan waktu tempuh. Dalam model deterministik, jika jarak X diketahui maka secara pasti kita dapat memprediksi lama waktu tempuh yang diperlukan (Y) tanpa adanya kekeliruan atau galat (*error*).

Dalam realitas kehidupan 'kondisi ideal' jarang dijumpai. Meskipun jarak dari rumah Anda ke kampus diketahui 20 km, tetapi Anda tidak dapat

memastikan lama waktu tempuh dari rumah ke kampus karena Anda akan sulit melakukan perjalanan dengan kecepatan konstan. Jika Anda melakukan perjalanan setiap harinya dengan kecepatan rata-rata 40 km/jam maka waktu tempuh yang Anda perlukan rata-rata setengah jam. Waktu tempuh Anda dari rumah ke kampus dari hari ke hari akan bervariasi, bisa lebih atau kurang dari setengah jam.

Kita dapat memodifikasi model deterministik (1.1) untuk memodelkan waktu tempuh Anda dari rumah ke kampus sebagai fungsi dari jarak. Caranya adalah dengan memasukkan suku galat acak (*random error*) ke dalam model (1.1), sehingga modelnya menjadi:

$$Y = 0{,}025\ X + \text{Galat Acak}$$

Model ini dinamakan **model probabilistik**. Suku galat acak dipakai untuk memperhitungkan adanya variasi nilai-nilai Y untuk nilai X yang tetap. Suku $0{,}025X$ merupakan komponen deterministik. Dalam banyak situasi model probabilistik lebih sesuai untuk digunakan.

1.2 Model-Model Regresi

Dalam buku ini kita akan mempelajari model probabilistik yang dinamakan model regresi. Model regresi yang paling sederhana adalah model *regresi linier sederhana* dengan bentuk persamaan

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (1.2)$$

Komponen deterministik pada model (1.2) adalah $\beta_0 + \beta_1 X$ jika X dianggap bukan variabel acak. Variabel Y dinamakan variabel dependen, variabel X dinamakan variabel independen, β_0 (beta nol) dan β_1 (beta satu) adalah parameter-parameter yang nilainya tidak diketahui, dan ε (epsilon) adalah galat acak (*random error*). Komponen deterministik pada model (1.2) adalah $\beta_0 + \beta_1 X$ *jika X dianggap bukan variabel acak".* Ada istilah-istilah lain untuk pasangan variabel independen-variabel dependen, yakni variabel bebas-variabel terikat, *explanatory variable-explained variable*, *predictor-predictand*, *regressor-regressand*, *stimulus-response*, *exogenous-endogenous*, *covariate-outcome*; atau *control variable-*

controlled variable. Untuk selanjutnya dalam buku ini akan digunakan istilah variabel independen dan variabel dependen.

Model (1.2) disebut **regresi linier** karena Y merupakan fungsi linier dari X, yakni variabel X berpangkat 1. Pada model ini Y juga merupakan fungsi linier dari parameter karena β_0 dan β_1 masing-masing juga berpangkat 1. Secara umum model yang linier dalam parameter dinamakan **model linier**. Jadi, model (1.2) sekaligus merupakan *model regresi linier* dan juga *model linier*. Kita akan membahas secara detail model regresi linier sederhana di Bab 2.

Model regresi linier dapat melibatkan lebih dari satu variabel independen. Secara umum jika terdapat k variabel independen, yakni X_1, X_2, ..., X_k, maka model regresi liniernya adalah

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + \varepsilon \qquad (1.3)$$

Model (1.3) dinamakan model regresi linier ganda (*multiple*). Di sini β_0, β_0, ..., β_k adalah parameter dan ε adalah galat acak. Perhatikan bahwa model (1.3) juga merupakan model linier. Kita secara khusus akan membahas regresi linier ganda dengan dua dan tiga variabel independen di Bab 3. Di Bab 4 akan dibahas model regresi linier secara umum dengan menggunakan notasi matriks.

Model regresi dengan persamaan berbentuk

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon \qquad (1.4)$$

bukan model regresi linier karena memuat X^2, tetapi merupakan *model linier* karena linier dalam parameter. Jika dimisalkan $X_1 = X$ dan $X_2 = X^2$ maka model (1.4) dapat ditulis sebagai

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

yang berbentuk seperti model regresi linier ganda dengan dua variabel independen. Model seperti ini dan beberapa variasi yang lain akan dibahas di Bab 5.

Terdapat juga model-model regresi yang tidak linier, baik dalam variabel independen maupun dalam parameter, misalnya:

$$Y = \beta_0 e^{\beta_1 X} + \varepsilon \,.$$

Di sini variabel independen X dan parameter β_1 sebagai pangkat. Model

$$Y = \frac{1}{\beta_0 + \beta_1 X} + \varepsilon$$

juga tidak linier baik dalam variabel independen maupun dalam parameter. Di dalam buku ini tidak dibahas model regresi yang tidak linier dalam parameter.

Bab 2
Regresi Linier Sederhana

2.1 Pendahuluan

Model regresi linier sederhana adalah model probabilistik yang menyatakan hubungan linier antara dua variabel di mana salah satu variabel dianggap memengaruhi variabel yang lain. Variabel yang memengaruhi dinamakan variabel independen dan variabel yang dipengaruhi dinamakan variabel dependen. Sebagai contoh, mungkin seorang peneliti tertarik untuk menyelidiki pengaruh (hubungan) linier dari *intelegency quotient* (IQ) terhadap hasil belajar statistika mahasiswa. Di sini IQ adalah variabel independen, sedangkan hasil belajar statistika adalah variabel dependen. Masih banyak contoh yang dapat dimodelkan dengan regresi linier sederhana, misalnya hubungan antara motivasi dan kinerja pegawai, hubungan antara usia dan tinggi badan manusia, hubungan antara pendapatan dan pengeluaran rumah tangga, dan lain-lain.

Di bab ini akan dibahas secara detail model regresi linier sederhana. Pada Bagian 2.2 dibahas model matematika untuk regresi linier sederhana. Bagian 2.3 dan 2.4 berturut-turut berisi tentang estimasi dan uji hipotesis tentang parameter. Pada Bagian 2.5 dibahas interval konfidensi. Pada bagian ini juga dijelaskan kaitan antara interval konfidensi dan uji hipotesis. Bagian 2.6 berisi pembahasan tentang koefisien determinasi dan Bagian 2.7 membahas penggunaan model untuk estimasi dan prediksi. Bagian 2.8 berisi hubungan antara regresi dan korelasi.

2.2 Model Regresi Linier Sederhana

Model probabilistik untuk regresi linier sederhana adalah:

$$Y = \beta_0 + \beta_1 X + \varepsilon \qquad (2.1)$$

di mana X adalah variabel independen, Y adalah variabel dependen, β_0 dan β_1 adalah parameter-parameter yang nilainya tidak diketahui yang dinamakan koefisien regresi, dan ε adalah kekeliruan atau galat acak (*random error*). *Di sini variabel independen X* ***diasumsikan bukan variabel acak****, dapat diobservasi atau diukur dengan kekeliruan yang dapat diabaikan, dan variasi dalam X dianggap dapat diabaikan dibanding dengan range dari X*. Sebagai konsekuensi dari adanya suku galat acak ε maka variabel dependen Y juga merupakan variabel acak.

Galat acak ε mempunyai peranan yang sangat penting dalam analisis regresi. Galat acak ε digunakan untuk memodelkan variasi nilai-nilai Y untuk nilai X yang tetap. Sebagai contoh, dari 10 mahasiswa dengan tingkat IQ yang sama (X), jika diuji maka hasil belajarnya (Y) belum tentu sama, tetapi pada umumnya akan ada variasi. Variasi ini mungkin karena ada faktor selain IQ yang memengaruhi hasil belajar. Karena kita hanya fokus pada pengaruh X terhadap Y, maka akan selalu diasumsikan bahwa *mean* (harga harapan atau ekspektasi) galat acak ε sama dengan 0, ditulis $E(\varepsilon) = 0$. Ini berarti bahwa pengaruh semua faktor di luar X *mean*-nya dianggap sama dengan 0. Asumsi ini kiranya beralasan untuk mendapatkan model regresi linier sederhana yang baik.

Dengan asumsi bahwa *mean* galat acak sama dengan nol, maka *mean* variabel dependen Y dinotasikan dengan $E(Y)$ adalah:

$$E(Y) = \beta_0 + \beta_1 X. \qquad (2.2)$$

Dari rumus ini terlihat bahwa *mean* dari Y hanya dipengaruhi oleh X, parameter β_0 dan β_1, dan tidak dipengaruhi oleh faktor lain. Persamaan (2.2) merupakan persamaan garis lurus dengan gradien (kemiringan) β_1 yang memotong sumbu vertikal di β_0. Parameter β_0 dinamakan *intercept* dan parameter β_1 menyatakan perubahan pada *mean* $E(Y)$ untuk setiap kenaikan satu satuan dalam X.

Jika (X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n) adalah sampel dari pasangan variabel independen X dan variabel dependen Y yang memenuhi persamaan (2.1) maka,

$$
\begin{aligned}
Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\
Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\
&\vdots \\
Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n
\end{aligned}
\qquad (2.3)
$$

Di sini, terdapat n galat acak ε_1, ε_2, ..., ε_n. Semua galat acak ini diasumsikan memiliki *mean* 0.

Kita juga akan selalu asumsikan bahwa galat acak ε memiliki variansi konstan σ^2 (sigma kuadrat), ditulis Var(ε) = σ^2. Untuk nilai-nilai X yang berbeda galat-galat acaknya dianggap mempunyai variansi yang sama, yakni semua galat acak ε_1, ε_2, ..., ε_n pada rumus (2.3) diasumsikan semuanya memiliki variansi σ^2. Sebagai akibatnya, Var(Y_i) = σ^2 untuk setiap i = 1, 2, ..., n. Asumsi ini dikenal dengan asumsi homogenitas atau dalam analisis regresi sering disebut homoskedastisitas (*homoscedasticity*).

Dalam praktik mungkin saja untuk nilai X yang berbeda variasi nilai-nilai Y juga berbeda. Sebagai contoh, variasi hasil belajar mahasiswa dengan tingkat IQ 90 mungkin berbeda dengan variasi hasil belajar mahasiswa dengan tingkat IQ 130, tetapi yang dibahas di sini semua variansi galat acak diasumsikan sama. Tujuannya adalah untuk memudahkan dan menyederhanakan analisis.

Asumsi lain yang nantinya akan digunakan adalah bahwa galat-galat acak ε_1, ε_2, ..., ε_n tidak berkorelasi. Ada juga yang mengasumsikan bahwa galat-galat acak saling independen. Ini dua hal yang sedikit berbeda. Dalam teori statistik, jika galat-galat acak diasumsikan saling independen, maka galat-galat acak tersebut pasti tidak berkorelasi, tetapi jika galat-galat acak tidak berkorelasi belum tentu galat-galat acak tersebut saling independen. Galat-galat acak yang tidak berkorelasi juga akan independen jika galat-galat acak tersebut berdistribusi normal.

Selanjutnya galat acak ε akan diasumsikan berdistribusi normal dengan *mean* 0 dan variansi konstan σ^2 untuk sembarang nilai variabel independen X. Dengan demikian, galat-galat acak ε_1, ε_2, ..., ε_n semuanya diasumsikan memiliki distribusi normal dengan *mean* 0 dan variansi sama

σ^2. Asumsi normalitas sangat penting dalam analisis regresi, khususnya dalam uji hipotesis atau untuk membentuk interval konfidensi parameter. Dengan mengasumsikan galat acak ε berdistribusi normal, akan dapat dikenali distribusi-distribusi statistik untuk uji hipotesis, dan selanjutnya kita dapat memanfaatkan tabel-tabel yang tersedia untuk membuat kesimpulan dalam uji hipotesis tentang parameter. Perlu ditegaskan di sini bahwa dalam analisis regresi yang diuji normalitasnya adalah galat acaknya, bukan variabel dependen atau bahkan variabel independennya. Hal ini karena mungkin saja galat-galat acak berdistribusi normal, tetapi data variabel dependen bukan dari distribusi normal.

2.3 Mengestimasi Parameter

Salah satu hal yang sangat penting dalam analisis regresi adalah mengestimasi parameter β_0, β_1, dan σ^2.

2.3.1 Mengestimasi β_0 dan β_1

Anggap kita telah memiliki realisasi data sampel pasangan variabel independen dan variabel dependen (X_1,Y_1), (X_2,Y_2), ..., (X_n,Y_n). Pasangan (X_1,Y_1) berarti bahwa dari responden pertama telah diperoleh data variabel independen X_1 dan variabel dependen Y_1, pasangan (X_2,Y_2) berarti bahwa dari responden kedua telah diperoleh data variabel independen X_2 dan variabel dependen Y_2, dan seterusnya. Dengan data sampel ini selanjutnya dapat diestimasi nilai-nilai parameter β_0 dan β_1.

Ada beberapa metode untuk mendapatkan estimator atau penduga untuk β_0 dan β_1. Dua di antara metode yang terkenal adalah metode kuadrat terkecil biasa (*ordinary least square*) dan metode kemungkinan maksimum (*maximum likelihood*).

Metode Kuadrat Terkecil

Untuk sekedar mengestimasi nilai-nilai parameter β_0 dan β_1 dengan metode kuadrat terkecil galat-galat acak tidak perlu diasumsikan bahwa memiliki *mean* 0, variansi konstan, tidak berkorelasi, atau berdistribusi

normal. Perhatikan bahwa dari persamaan (2.3) jika tersedia sampel dari n responden maka kita memiliki n galat acak, yaitu:

$$\begin{aligned} \varepsilon_1 &= Y_1 - \beta_0 + \beta_1 X_1 \\ \varepsilon_n &= Y_2 - \beta_0 + \beta_1 X_2 \\ &\vdots \\ \varepsilon_n &= Y_n - \beta_0 + \beta_1 X_n \end{aligned} \qquad (2.4)$$

Model yang ’terbaik’ adalah model yang memiliki nilai galat yang terkecil. Karena ada n nilai galat acak, yakni ε_1, ε_2, ..., ε_n, maka kita perlu memperhitungkan seluruh nilai galat tersebut. Sekilas mungkin kita memikirkan bahwa model yang terbaik adalah model yang jumlah seluruh galatnya nol. Kriteria ini kurang baik karena jika beberapa galat positif dan yang lain negatif, maka jumlahnya mungkin mendekati atau sama dengan 0, dan bukan berarti bahwa modelnya sudah baik. Ide yang lebih baik adalah mengkuadratkan seluruh galat (sehingga nilainya selalu positif) dan kemudian menjumlahkannya sehingga diperoleh jumlah kuadrat galat yang dapat dituliskan sebagai:

$$\sum_{i=1}^{n} \varepsilon_i^2$$

Jumlah kuadrat galat ini kemudian dibuat sekecil mungkin. Nilai β_0 dan β_1 yang membuat jumlah kuadrat galat ini bernilai minimum merupakan estimator untuk β_0 dan β_1, yang selanjutnya masing-masing akan dinotasikan dengan b_0 dan b_1. Estimator ini dinamakan estimator kuadrat terkecil (*least squared estimator*). Dapat dibuktikan secara matematis bahwa rumus untuk b_1 adalah:

$$b_1 = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2} \qquad (2.5)$$

sedangkan rumus untuk b_0 adalah

$$b_0 = \frac{\sum_{i=1}^{n} Y_i - b_1 \sum_{i=1}^{n} X_i}{n} \qquad (2.6)$$

atau

$$b_0 = \overline{Y} - b_1 \overline{X} \qquad (2.7)$$

atau

$$b_0 = \frac{\sum_{i=1}^{n} X_i^2 \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} X_i Y_i}{n \sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i \right)^2} \qquad (2.8)$$

Ketiga rumus b_0 di atas ekuivalen, yakni akan memberikan hasil yang sama jika dipakai untuk menghitung b_0. Untuk menghitung b_0 dan b_1 penulis menyarankan pertama-tama menghitung b_1 dengan menggunakan rumus (2.5) dan selanjutnya menghitung b_0 dengan rumus (2.6) atau (2.7). Rumus untuk b_1 juga dapat disajikan dalam bentuk:

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad (2.9)$$

di mana

$$x_i = X_i - \overline{X} \text{ dan } y_i = Y_i - \overline{Y} .$$

Perhatikan bahwa huruf besar dan huruf kecil pada rumus di atas mempunyai arti yang berbeda. Jika kita menggunakan rumus (2.9) untuk menghitung b_1 maka sebaiknya menggunakan rumus (2.7) untuk menghitung b_0.

Rumus (2.9) juga dapat dituliskan dalam bentuk lain, yakni:

$$b_1 = \frac{J_{XY}}{J_{XX}} \tag{2.10}$$

di mana

$$J_{XY} = \sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y}) = \sum_{i=1}^{n}(X_i - \overline{X})Y_i = \sum_{i=1}^{n} X_i Y_i - \frac{\sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n} \tag{2.11}$$

dan

$$J_{XX} = \sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n} \tag{2.12}$$

Perhitungan secara manual untuk b_0 dan b_1 akan lebih mudah jika digunakan bantuan tabel, lihat Contoh 2.1.

Di awal telah disebutkan bahwa untuk *sekedar* mengestimasi nilai-nilai parameter β_0 dan β_1 dengan metode kuadrat terkecil, maka asumsi-asumsi yang telah disebutkan terdahulu tidak diperlukan, tetapi kita *tidak dapat menilai* baik tidaknya estimator yang diperoleh. Jika galat-galat acak ε_1, ε_2, ..., ε_n diasumsikan semuanya memiliki *mean* 0 dan variansi σ^2 dan tidak korelasi (tidak perlu asumsi normalitas), maka estimator-estimator b_0 dan b_1 yang diperoleh memiliki sifat-sifat yang baik, yakni tidak bias dan memiliki variansi terkecil di antara estimator-estimator linier lainnya, atau dikenal dengan *Best Linear Unbiased Estimator* (BLUE).

Metode Maksimum Likelihood

Untuk mengestimasi nilai parameter β_0 dan β_1 dengan metode maksimum *likelihood* diperlukan asumsi bahwa variabel-variabel galat ε_1, ε_2, ..., ε_n, semuanya memiliki *mean* 0, variansi σ^2, tidak berkorelasi/ independen, dan berdistribusi normal. Penjelasan metode ini terlalu matematis sehingga kurang bijaksana untuk diuraikan di sini. Hal yang

sangat penting untuk diketahui oleh para pembaca adalah bahwa dalam regresi linier sederhana (dan juga pada regresi linier ganda yang akan dibahas di bab selanjutnya) metode maksimum *likelihood* memberikan hasil estimator yang sama dengan estimator yang diperoleh dengan metode kuadrat terkecil.

Estimator yang diperoleh dengan metode maksimum *likelihood* mempunyai sifat-sifat yang baik. Untungnya dalam analisis regresi linier estimator, untuk β_0 dan β_1 yang diperoleh dengan metode maksimum *likelihood* dan metode maksimum *likelihood* sama sehingga kita cukup puas menggunakan estimator yang diperoleh dengan metode kuadrat terkecil. Secara umum, metode maksimum *likelihood* dan metode kuadrat terkecil tidak selalu memberikan hasil yang sama, misalnya jika distribusi galat acak tidak normal.

Setelah nilai b_0 dan b_1 diperoleh maka estimasi hubungan antara variabel independen X dan variabel dependen Y dapat dituliskan sebagai:

$$\hat{Y} = b_0 + b_1 X \tag{2.13}$$

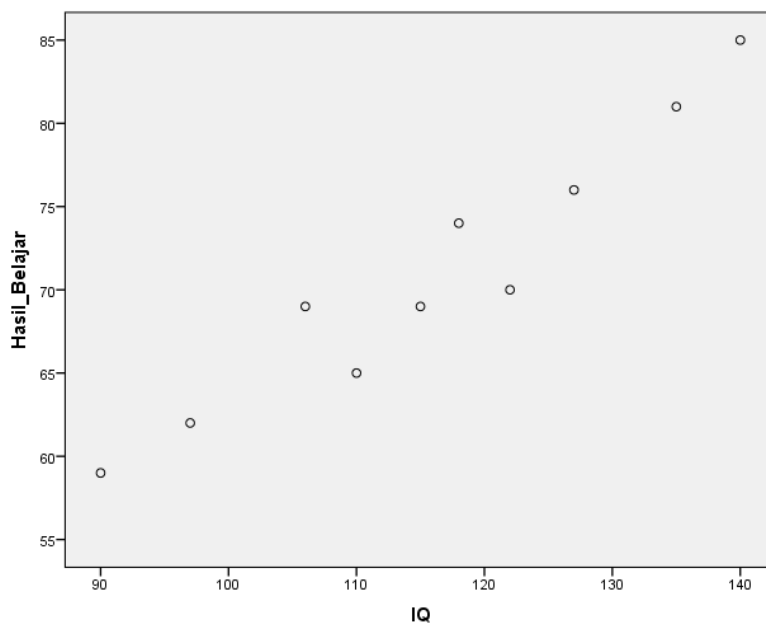
Persamaan ini dinamakan **persamaan garis regresi.**

Contoh 2.1 (Mengestimasi β_0 dan β_1)

Anggap seorang peneliti meyakini ada hubungan atau pengaruh linier dari tingkat IQ terhadap hasil belajar mahasiswa. Tingkat IQ dianggap sebagai variabel independen (X) dan hasil belajar mahasiswa dianggap sebagai variabel dependen (Y). Anggap data sampel yang tersedia (data simulasi) dari responden sebanyak 10 mahasiswa adalah sebagai berikut:

Tabel 2.1 Data Simulasi Tingkat IQ dan Hasil Belajar Mahasiswa

Nomor Responden	Tingkat IQ (X)	Hasil Belajar Mahasiswa (Y)
1	90	59
2	97	62
3	106	69
4	110	65
5	115	69
6	118	74
7	122	70
8	127	76
9	135	81
10	140	85

Diagram pencar (*scatter plot*) untuk data di atas terlihat pada Gambar 2.1.

Gambar 2.1 Diagram Pencar Data Tingkat IQ dan Hasil Belajar Mahasiswa